Capítulo 5

Matrizes e Sistemas lineares

Neste capítulo estudaremos alguns métodos para calcular a solução de sistemas de equações lineares. Apenas nos preocuparemos com sistemas quadrados, isto é, aqueles em que o número de equações é igual ao número de incógnitas. Supõe-se que as noções básicas de álgebra matricial, como adição e multiplicação de matrizes, matriz inversa e identidade, determinante de uma matriz etc., sejam conhecidas do leitor.

5.1 Introdução

Um sistema de equações algébricas de ordem n, que é um conjunto de n equações com n incógnitas,

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n \end{cases}$$

pode ser representado através de uma equação matricial

$$Ax = b\,,$$

onde

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \text{ é matrix dos coeficientes,}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ é o vetor colunar das incógnitas,}$$

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \text{ é o vetor dos termos independentes.}$$

Em todo o texto, salvo menção em contrário, sempre indicaremos um sistema linear genérico de ordem n por $Ax = b$. Para facilidade de notação usaremos indistintamente

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ ou } x = (x_1, \ldots, x_n)\,.$$

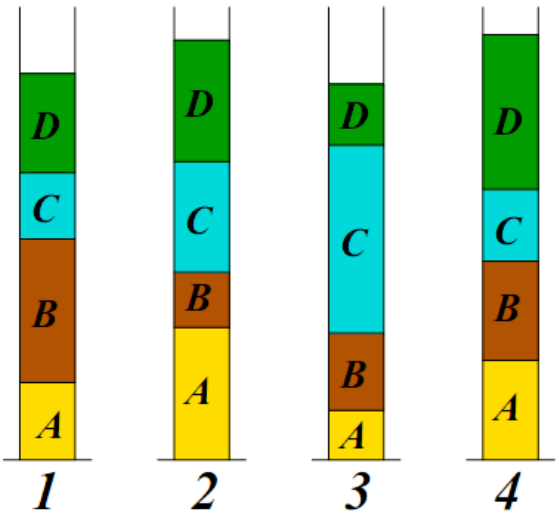
5.2 Exemplos de Aplicação

5.2.1 Provetas [1]

Considere o seguinte problema: quatro tipos de materiais particulados estão distribuídos por quatro provetas, e em cada proveta os materiais são dispostos em camadas, não misturadas, de modo que seja possível medir facilmente o volume de cada material em cada uma delas. Dado que possamos medir a massa total de cada proveta, e que saibamos a massa da proveta vazia, queremos calcular a densidade de cada um dos materiais.

Para colocar o problema em termos matemáticos, chamemos os materiais de A, B, C e D, e suas densidades respectivas de ρ_A, ρ_B, ρ_C e ρ_D. Essas são as *incógnitas* do problema, números que queremos descobrir.

Entre os dados disponíveis para resolvê-lo estão a massa conjunta dos quatro materiais em cada uma das provetas (numeradas de 1 a 4), que chamaremos de m_1, m_2, m_3 e m_4, já descontada a tara das provetas.



Além disso, temos o volume de cada um dos materiais em cada uma das provetas. Chamaremos de v_{1A}, v_{1B}, v_{1C} e v_{1D} o volume dos materiais A, B, C e D na Proveta 1, v_{2A}, v_{2B}, v_{2C} e v_{2D} o volume dos materiais A, B, C e D na Proveta 2, e assim por diante.

Como a densidade é a razão entre massa e volume, a massa do material A na Proveta 1 é $v_1 \times \rho_A$. Estendendo esse raciocínio para os demais materiais, obtemos que a massa total m_1 contida na Proveta 1 é

$$v_{1A} \times \rho_A + v_{1B} \times \rho_B + v_{1C} \times \rho_C + v_{1D} \times \rho_D \,.$$

Considerando as quatro provetas, obteremos quatro equações:

$$\begin{cases} v_{1A} \times \rho_A + v_{1B} \times \rho_B + v_{1C} \times \rho_C + v_{1D} \times \rho_D & = m_1 \\ v_{2A} \times \rho_A + v_{2B} \times \rho_B + v_{2C} \times \rho_C + v_{2D} \times \rho_D & = m_2 \\ v_{3A} \times \rho_A + v_{3B} \times \rho_B + v_{3C} \times \rho_C + v_{3D} \times \rho_D & = m_3 \\ v_{4A} \times \rho_A + v_{4B} \times \rho_B + v_{4C} \times \rho_C + v_{4D} \times \rho_D & = m_4 \end{cases}$$

Trata-se de um sistema linear de quatro equações e quatro incógnitas.

[1] Extraído de Asano & Coli 2009

Uma possível aplicação em geologia seria a seguinte. Uma sonda faz o papel das provetas, e uma coluna de material é retirada, contendo materiais diferentes dispostos em camadas (pode ser até uma sonda coletando material congelado). A sonda permitiria medir a dimensão de cada camada, mas não poderíamos desmanchar a coluna para medir a densidade de cada material isoladamente, sob o risco de alterar a compactação.

5.2.2 Resolução do Círculo

Vamos agora concluir o exemplo iniciado no Capítulo 2. Nosso problema era o seguinte: dadas as coordenadas de três pontos quaisquer, (x_1, y_1), (x_2, y_2), (x_3, y_3), resolver a equação do círculo que passa por estes três pontos,

$$(x-a)^2 + (y-b)^2 = r^2\,,$$

de forma a determinar o centro do círculo (a, b) e seu raio, r. Temos três incógnitas, de forma que são necessárias três equações para resolver o problema. São elas

$$\begin{cases} (x_1-a)^2 + (y_1-b)^2 &= r^2 \\ (x_2-a)^2 + (y_2-b)^2 &= r^2 \\ (x_3-a)^2 + (y_3-b)^2 &= r^2\,. \end{cases}$$

Vamos inicialmente manipular a primeira equação. Expandindo os termos quadráticos obtemos

$$x_1^2 + a^2 - 2ax_1 + y_1^2 + b^2 - 2by_1 - r^2 = 0\,.$$

Definindo

$$k \equiv a^2 + b^2 - r^2$$

obtemos

$$2x_1a + 2y_1b - k = x_1^2 + y_1^2\,.$$

Manipulando as demais equações da mesma forma, obtemos o seguinte sistema de equações lineares

$$\begin{cases} 2x_1a + 2y_1b - k &= x_1^2 + y_1^2 \\ 2x_2a + 2y_2b - k &= x_2^2 + y_2^2 \\ 2x_3a + 2y_3b - k &= x_3^2 + y_3^2\,. \end{cases}$$

que, escrito em forma matricial, fica

$$\begin{bmatrix} 2x_1 & 2y_1 & -1 \\ 2x_2 & 2y_2 & -1 \\ 2x_3 & 2y_3 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ k \end{bmatrix} = \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ x_3^2 + y_3^2 \end{bmatrix}\,. \tag{5.1}$$

O problema resume-se, agora, em resolver o sistema acima para obter a, b e k.

5.2.3 Calculando as populações do H em uma região H II

5.3 Método de Cramer

Um método para resolver sistemas lineares, talvez já conhecido do leitor, é o método de Cramer. Nele a solução do sistema $Ax = b$ é dada por

$$x_i = \frac{\det(A_i)}{\det(A)}, i = 1, 2, \ldots, n$$

onde $\det(A)$ é o determinante da matriz A, e A_i é a matriz obtida de A substituindo a sua i-ésima coluna pelo vetor b dos termos independentes.

O determinante de uma matriz A de ordem n pode ser calculado através do desenvolvimento por linhas (regra de Laplace):

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

onde i é o índice de uma linha qualquer e A_{ij} é a matriz obtida de A retirando-se a i-ésima linha e a j-ésima coluna.

Observe que se $\det(A) \neq 0$ então o sistema $Ax = b$ tem uma única solução. Se $\det(A) = 0$ então podem ocorrer dois casos:

1. o sistema não possui solução (sistema inconsistente);

2. o sistema possui infinitas soluções (sistema indeterminado).

Por exemplo, no caso de um sistema linear de ordem 2, cada equação representa uma reta. Resolver o sistema significa determinar a intersecção das duas retas. Se as duas retas forem coincidentes, então há infinitos pontos de interseção. Se forem paralelas, não há nenhum ponto de interseção. Neste texto nos preocuparemos com sistemas lineares que tenham uma única solução.

Uma das propriedades do determinante é que se uma das linhas da matriz for uma combinação linear de outra (ou outras), então o determinante será zero. Por exemplo, a matriz abaixo tem determinante zero

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}.$$

Sistemas que possuem equações que são combinações lineares são ditos degenerados ou singulares. Como vimos acima, eles podem ser ou inconsistentes ou indeterminados.

Sistemas não singulares (em que o $\det(A) \neq 0$) possuem sempre uma solução. Entretanto, duas questões numéricas podem impedir que a solução seja obtida

1. embora não sejam combinações lineares exatas de outras, algumas equações podem ser tão próximas a combinações lineares que erros de arredondamento as tornem linearmente dependentes em algum estágio da solução. Neste caso, o procedimento numérico irá falhar.

2. Erros de arredondamento cumulativo podem impedir que a solução seja obtida.

Ao longo deste capítulo discutiremos formas de lidar com estas duas questões.

A utilização do método de Cramer para resolver sistemas lineares pode ser inviável, pois o número de operações aritméticas que devem ser efetuadas aumenta consideravelmente com um pequeno aumento na ordem do sistema.

Para estimar o número de operações necessárias para a regra de Cramer, vamos considerar o caso de um sistema com $n = 20$. Para resolvê-lo, precisamos calcular 21 determinantes de ordem 20. Mas, para calcular um determinante de ordem 20, usamos a regra de Laplace, que decompõe o determinante em uma soma envolvendo 20 determinantes de ordem 19. Se extrapolarmos o processo até chegarmos em determinantes de ordem 2, teremos que o número de operações aritméticas será da ordem de $21! \approx 5 \times 10^{19}$. Para um sistema de ordem n, temos que o número de operações será da ordem de $(n+1)!$.

Em um computador pessoal de 30 Gflops[2] estas 10^{20} operações levariam 3.3×10^9 s ou aproximadamente 100 anos! Na prática, a situação é ainda pior, pois estamos considerando apenas o tempo para efetuar as operações aritméticas, e não o acesso à memória.

No novo super-computador do IAG, que terá uma capacidade teórica de 20 Tflops, esta conta levaria "apenas" 57 dias. Embora útil para sistemas de ordem menor, o método de Cramer é impraticável para sistemas maiores, e outros métodos devem ser empregados neste caso. Outro aspecto negativo do método de Cramer é que como ele necessita de muitas operações aritméticas, ele potencialmente gerará mais erros de arredondamento.

Exercício 1: use a regra de Cramer para obter uma solução analítica para o problema do círculo (Eq. 5.1).

5.4 Tarefas da álgebra linear computacional

Há muito mais na álgebra linear do que resolver um único sistema de equações lineares. Abaixo listamos os principais tópicos abordados neste capítulo.

- Solução para a equação matricial $Ax = b$ para um vetor colunar desconhecido, x.
- Solução para mais de uma de uma equação matricial, $Ax_j = b_j$, para um conjunto de vetores x_j, $j = 0, 1, \ldots$, cada um correspondendo a um dado vetor de termos independentes, b_j. Nesta tarefa a simplificação chave é que a matriz A é mantida constante, enquanto que os termos independentes variam.
- Cálculo da matriz inversa A^{-1}, que obedece à equação matricial $AA^{-1} = I$, onde I é a matriz identidade.
- Cálculo do determinante de uma matriz quadrada A.
- Melhora iterativa da solução de um sistema.

5.5 Sistemas de acordo com as propriedades das matrizes

Tipicamente, podemos ter dois tipos de sistemas lineares, os sistemas *cheios* e *esparsos*. Nos sistemas cheios, todos, ou ao menos a grande maioria, dos elementos da matriz A é diferente de zero. Nos sistemas esparsos, uma parte importante dos elementos de A é nula. Um caso importante são sistemas com matrizes tridiagonais, como ilustrado na Figura 5.1.

Sistemas esparsos possuem soluções particulares e mais rápidas que os sistemas cheios. Vamos inicialmente estudar os métodos de solução para matrizes cheias.

5.6 Método da Eliminação de Gauss

5.6.1 Sobre o método

É o método mais simples para solução de um sistema de equações. O método de Gauss possui várias características que o tornam interessante, mesmo que haja métodos mais eficientes.

Uma característica interessante do método é que quando aplicado para resolver um conjunto de equações lineares, a eliminação de Gauss produz tanto a solução das equações

[2]Flops significa número de operações de ponto flutuante por segundo.

Figura 5.1: Exemplos de matrizes esparsas.

(para um ou mais vetores de termos independentes) quanto a inversa da matriz A (esta última é obtida quando empregamos uma variante do método, chamada de método de Gauss-Jordan, seção 5.7). Uma de suas características mais importantes é que o método é tão estável quanto qualquer outro método direto (direto, aqui, é usado em contraposição aos métodos iterativos mostrados no fim do capítulo), desde que seja empregado o pivotamento (seções 5.6.4 e 5.7.2)

Algumas deficiências do método são

1. se a matriz inversa não for desejada, o método de Gauss é tipicamente 3 vezes mais lento que a melhor alternativa disponível (decomposição LU, seção 5.10).

2. quanto o empregamos para mais de uma equação matricial ($Ax_j = b_j$), todos os vetores de termos independentes devem ser armazenados na memória e manipulados simultaneamente.

A deficiência 1) acima pode suscitar questionamentos, afinal, se temos a matriz inversa, podemos calcular as incógnitas de um sistema $Ax_j = b_j$ através de:

$$x_j = A^{-1}b_j \, .$$

Isto realmente funciona, mas este procedimento resulta em uma resposta muito suscetível a erros de arredondamento, e deve ser evitado.

5.6.2 Procedimento

Vamos ilustrar o procedimento do método de eliminação de Gauss com um exemplo simples. O objetivo consiste em transformar o sistema $Ax = b$ em um sistema triangular equivalente. Para isso, usarmos a seguinte propriedade da Álgebra Linear.

Propriedade: A solução de um sistema linear não se altera se subtrairmos de uma equação outra equação do sistema multiplicada por uma constante.

Considere o seguinte sistema de equações:

$$\begin{cases} 2x + \ y + \ z = \ 7 \\ 4x + 4y + 3z = 21 \\ 6x + 7y + 4z = 32 \end{cases}$$

Multiplicando a primeira equação por (-2) e somando na segunda, e multiplicando a primeira equação por (-3) e somando na terceira temos

$$\begin{cases} 2x + \ y + z = \ 7 \\ \quad\ 2y + z = \ 7 \\ \quad\ 4y + z = 11 \end{cases}$$

Multiplicando a segunda equação por (-2) e somando na terceira temos

$$\begin{cases} 2x + \ y + z = \ 7 \\ \quad\ 2y + z = \ 7 \\ \quad\ \ \ -z = -3 \end{cases}$$

Da terceira equação temos $-z = -3 \Rightarrow \boxed{z = 3}$.
Substituindo na segunda equação temos $2y + 3 = 7 \Rightarrow \boxed{y = 2}$.
Substituindo na primeira equação temos $2x + 2 + 3 = 7 \Rightarrow \boxed{x = 1}$.

Vemos que o método de Gauss é uma forma sistemática de triangularizar um sistema linear. A solução é obtida em dois passos:

1. Eliminação (*forward elimination*): triangularização propriamente dita.
2. Substituição (*back substitution*): obtenção da solução final (vetor x).

Se usarmos a notação matricial, estamos resolvendo a equação

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 4 & 3 \\ 6 & 7 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \\ 32 \end{bmatrix}$$

transformando-a em

$$\underbrace{\begin{bmatrix} 2 & 1 & 1 \\ & 2 & 1 \\ & & -1 \end{bmatrix}}_{\substack{\text{matriz} \\ \text{triangular superior}}} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ -3 \end{bmatrix}.$$

Entretanto, podemos trabalhar somente com os números sem escrever as equações. Para tanto é conveniente escrever a chamada *matriz aumentada*

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 4 & 4 & 3 & 21 \\ 6 & 7 & 4 & 32 \end{array}\right].$$

Como antes multiplicamos a primeira equação por 2 e subtraímos da segunda; multiplicamos a primeira equação por 3 e subtraímos da terceira:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 0 & 2 & 1 & 7 \\ 0 & 4 & 1 & 11 \end{array}\right].$$

Então multiplicamos a segunda equação por 2 e subtraímos da terceira

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 0 & 2 & 1 & 7 \\ 0 & 0 & -1 & -3 \end{array}\right].$$

5.6.3 Estimativa do número de operações realizadas

Vamos estimar o número de operações realizadas na obtenção da solução x. Estimaremos separadamente o número de operações feitas durante a eliminação e a substituição.

1) Processo de eliminação

Para estimar o número de operações realizadas durante a triangulação da matriz, calcularemos quantas adições e multiplicações são necessárias em cada etapa do processo. Por exemplo, para eliminarmos a primeira coluna, temos $(n-1)$ linhas onde para cada uma delas são calculadas $n+1$ multiplicações e n adições.

eliminação da:	multiplicações	adições
1^{a} coluna	$(n-1)(n+1)$	$(n-1)n$
2^{a} coluna	$(n-2)n$	$(n-2)(n-1)$
$\vdots$	$\vdots$	$\vdots$
$(n-1)^{\text{a}}$ coluna	$(1)(3)$	$(2)(1)$
Total	$\sum_{i=1}^{n-1} i(i+2)$	$\sum_{i=1}^{n-1}(i+1)i$

O total de multiplicações é

$$\sum_{i=1}^{n-1} i(i+2) = \sum_{i=1}^{n-1} i^2 + 2\sum_{i=1}^{n-1} i\,.$$

Avaliando cada uma das somatórias

$$\sum_{i=1}^{n-1} i^2 = \sum_{i=1}^{n} i^2 - n^2 = \frac{n(n+1)(n+2)}{6} - n^2 = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}$$
$$\sum_{i=1}^{n-1} i = \sum_{i=1}^{n} i - n = \frac{n(n+1)}{2} - n = \frac{n^2}{2} - \frac{n}{2}\,,$$

que implica

$$\sum_{i=1}^{n-1} i^2 + 2\sum_{i=1}^{n-1} i = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} + 2\left(\frac{n^2}{2} - \frac{n}{2}\right) = \frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6}\,.$$

O número total de adições pode ser obtido de forma análoga:

$$\sum_{i=1}^{n-1}(i+1)i = \frac{n^3}{3} - \frac{n}{3}\,.$$

Obtemos, assim, que o número total de operações de ponto flutuante para o processo de eliminação é

$$N_{\text{elim}} = \frac{2n^3}{3} + \frac{n^2}{2} - \frac{7n}{6}\,.$$

Para um valor de n suficientemente grande, temos que os termos n^3 dominam nas expressões acima, de forma que o total de operações na eliminação será $\mathcal{O}(2n^3/3)$.

2) Processo de substituição

Vamos agora estimar quantas operações de ponto flutuante são feitas durante o cálculo da solução final a partir da matriz triangularizada (*back substitution*).

passo	multiplicações	adições
linha n	1	0
linha $n-1$	2	1
	$\vdots$	$\vdots$
linha 1	n	$n-1$
Total	$\sum_1^n i$	$\sum_1^{n-1} i$

Obtemos que o número de operações para esta fase

$$N_{\text{subst}} = n^2 \,.$$

Chegamos, assim, ao o número total de operações necessárias para resolver um sistema de ordem n pelo método de Gauss

$$N_{\text{Gauss}} = \frac{2n^3}{3} + \frac{3n^2}{2} - \frac{7n}{6} \,.$$

Concluímos que para valores altos de n o processo de eliminação necessita de um número muito maior de operações que a substituição e que, neste caso, o total de operações é

$$N_{\text{Gauss}} \approx \frac{2n^3}{3} \,.$$

Por exemplo, um sistema matricial de 20×20 implica em aproximadamente $2 \cdot 20^3/3 \approx 5 \cdot 10^3$ flop. Com um PC de 30 Gflops o problema será resolvido em

$$t = \frac{5 \cdot 10^3 \,\text{flop}}{30 \cdot 10^9 \,\text{flops}} \approx 2 \cdot 10^{-7} \,\text{s!}$$

Esta estimativa é muito otimista, pois consideramos que cada operação de ponto flutuante é efetuada em um ciclo da CPU. Isto é válido para adições, mas não para multiplicações, que tipicamente requerem da ordem de dez ciclos de CPU. Além disso não consideramos fatores como a perda de eficiência devido ao acesso à memória. De qualquer maneira, vemos que o método de Gauss é imensamente mais eficiente que o método de Cramer.

5.6.4 Pivotamento parcial

Seja o sistema

$$\begin{bmatrix} 10 & -7 & 0 \\ -3 & 2.099 & 6 \\ 5 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3.901 \\ 6 \end{bmatrix} \tag{5.2}$$

cuja solução é $x = [0, -1, 1]$.

Vamos considerar que nosso operador de ponto flutuante tenha apenas *5 algarismos significativos*, e vamos resolver o sistema acima pelo método de Gauss.

Multiplicando a 1ª equação por 0.3 e somando na 2ª; multiplicando a 1ª equação por -0.5 e somando na 3ª, obtemos

$$\left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & -0.001 & 6 & 6.001 \\ 0 & 2.5 & 5 & 2.5 \end{array}\right] . \tag{5.3}$$

Multiplicando a 2ª equação por $-2.5/-0.001 = 2500$ e somando na 3ª equação

$$\left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & -0.001 & 6 & 6.001 \\ 0 & 0 & 15005 & 15004 \end{array}\right].$$

Note que, devido à restrição de 5 algarismos significativos, tivemos que truncar as seguintes operações

$$\begin{aligned} 6.001 \times 2500 &= 15002.5\not{0} \\ 15002.5\not{0} + 2.5 &= 15004.5\not{0} = 15004\,. \end{aligned}$$

Ao efetuarmos a substituição obteremos

$$\begin{aligned} x_3' &= \frac{15004}{15005} = 0.99993 \\ x_2' &= \frac{6.001 - 6 \times 0.99993}{-0.001} = \frac{6.001 - 5.9995\not{8}}{-0.001} = -1.5 \\ x_1' &= \frac{7 + 7 \times (-1.5)}{10} = 7 - 10.510 = -0.35 \end{aligned}$$

Comparando este vetor $x' = (0.99993, -1.5, -0.35)$ obtido com o vetor $x = (1, -1, 0)$ solução, vemos o quão grande foi o erro gerado pela restrição de 5 algarismos significativos!

O que causou este problema? O primeiro elemento da linha que está sendo usada para eliminar os termos das demais é chamado de *pivô*. Na primeira etapa da eliminação acima (Eq. 5.3), o pivô, (-0.001), tornou-se muito pequeno em relação aos outros coeficientes, resultando num enorme multiplicador (2500) que fez aparecerem erros de arredondamento. Estes erros por sua vez são ampliados na fase de substituição, onde apareceram subtrações de números muito próximos divididas por números muito pequenos, o que amplifica enormemente o erro (por exemplo, veja o cálculo de x_2', acima).

Uma solução simples e eficiente para este problema é empregar *pivotamento parcial* no método de Gauss, que consiste em trocar linhas de forma que tenhamos sempre o maior valor absoluto possível para o pivô. Isto garantirá multiplicadores $\lesssim 1$ em módulo.

No exemplo acima, empregamos o pivotamento parcial já na primeira etapa

$$\left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & -0.001 & 6 & 6.001 \\ 0 & 2.5 & 5 & 2.5 \end{array}\right] \begin{array}{c} \text{pivotamento} \\ \longrightarrow \\ \text{parcial} \end{array} \left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & 2.5 & 5 & 2.5 \\ 0 & -0.001 & 6 & 6.001 \end{array}\right].$$

O multiplicador será $(-0.001)/(-2.5) = 0.0004$. Multiplicando a 2ª equação por este valor e somando na 3ª equação, obtemos a matriz estendida

$$\left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & 2.5 & 5 & 2.5 \\ 0 & 0 & 6.002 & 6.002 \end{array}\right],$$

que resulta na solução exata $x' = (1, -1, 0)$.

Uma regra importante a ser seguida: o *pivotamento parcial sempre deve ser empregado no método de Gauss!*

5.6.5 Solução simultânea de várias equações matriciais

Vimos na Seção 5.4 uma tarefa corriqueira da álgebra linear é resolver um conjunto de equações matriciais, $Ax_j = b_j$, $j = 1, \ldots, m$. Neste conjunto as equações matriciais compartilham a matriz A e possuem cada uma um dado dado vetor de termos independentes, b_j. Neste caso, em vez de fazer a mesma eliminação m vezes, podemos "guardar" a sequência de operações aplicadas na triangulação da matriz A para depois aplicar em b_j, $j = 1, \ldots, m$.

Por exemplo, seja a matriz

$$A = \begin{bmatrix} 2 & 6 & -2 \\ 1 & 3 & -4 \\ 3 & 6 & 9 \end{bmatrix}$$

e o vetor de pivotamento que contém o número da linha que foi pivotada,

$$p = \begin{bmatrix} \\ \\ \end{bmatrix}.$$

Com o pivotamento da 3ª linha, temos

$$\begin{bmatrix} 3 & 6 & 9 \\ 1 & 3 & -4 \\ 2 & 6 & -2 \end{bmatrix}; \quad \begin{bmatrix} 3 \\ \\ \end{bmatrix}.$$

Fazendo 1ª linha $\times (-1/3) +$ 2ª linha e 1ª linha $\times (-2/3) +$ 3ª linha,

$$\begin{bmatrix} 3 & 6 & 9 \\ \boxed{-1/3} & 1 & -7 \\ \boxed{-2/3} & 2 & -8 \end{bmatrix}; \quad \begin{bmatrix} 3 \\ \\ \end{bmatrix}.$$

Nesta operação, preservamos os *multiplicadores* que foram utilizados para eliminar os primeiros coeficientes das linhas que não eram o pivô. Para concluir a triangularização da matriz, novamente utilizamos o pivotamento da 3ª linha:

$$\begin{bmatrix} 3 & 6 & 9 \\ \boxed{-1/3} & 2 & -8 \\ \boxed{-2/3} & 1 & -7 \end{bmatrix}; \quad \begin{bmatrix} 3 \\ 3 \\ \end{bmatrix}.$$

Fazendo 2ª linha $\times (-1/2) +$ 3ª linha,

$$\begin{bmatrix} 3 & 6 & 9 \\ \boxed{-1/3} & 2 & -8 \\ \boxed{-2/3} & \boxed{-1/2} & -3 \end{bmatrix}; \quad \begin{bmatrix} 3 \\ 3 \\ \end{bmatrix}.$$

Para ilustrar como utilizar os multiplicadores armazenados e o vedor de pivotamento, vamos encontrar a solução para o vetor $b = [4, -7, 39]$.

- 1º passo: trocar a linha 1 com a linha $p(1) = 3 \rightarrow b = [39, -7, 4]$;
- 2º passo: multiplicar a 1ª linha por $\boxed{-1/3}$ e somar na 2ª linha; multiplicar a 1ª linha por $\boxed{-2/3}$ e somar na 3ª linha, $\rightarrow b = [39, -20, -22]$;

- 3º passo: trocar a linha 2 com a linha $p(2) = 3 \rightarrow b = [39, -22, -20]$;

- 4º passo: Multiplicar a 2ª linha por $\boxed{-1/2}$ e somar na 3ª linha, $b = [39, -22, -9]$.

Assim, o vetor x com a solução do sistema será dado por

$$\begin{bmatrix} 3 & 6 & 9 \\ 0 & 2 & -8 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 39 \\ -22 \\ -9 \end{bmatrix},$$

que pode ser facilmente resolvido por substituição

$$\begin{aligned} x_3 &= -9/(-3) = 3 \\ x_2 &= [-22 + 8 \times 3]/2 = 1 \\ x_1 &= [39 - 6 \times 1 - 9 \times 3]/3 = 2\,. \end{aligned}$$

.

O procedimento ilustrado pode ser repetido para um número arbitrário de vetores b_j. Uma sugestão para uma implementação eficiente do método de Gauss é fazer uma subrotina para a eliminação, que retorna a matriz triangularizada com os coeficientes de eliminação, segundo procedimento acima, e outra para a substituição.

Abaixo delineamos um possível algoritmo para implementar a eliminação de Gauss computacionalmente, mantendo os multiplicadores para uso posterior:

$$\begin{cases} \text{de } i = 1 \text{ até } n-1 \text{ faça} \\ \quad \text{determine o índice do pivoteamento } l \geq i \\ \quad \text{troque as linhas } i \text{ e } l, \text{ das colunas } i \text{ até } n \\ \quad \text{registre o } i\text{-ésimo pivotamento: } \; p(i) = l \\ \quad \text{laço} \begin{cases} \text{de } j = i+1 \text{ até } n \text{ faça} \\ \text{calcule o multiplicador para a linha } j \\ \text{guarde-o no lugar do elementos eliminado} \\ \text{adicione múltiplos da linha } l \text{ à linha } j \end{cases} \\ \text{fim do laço} \end{cases}$$

5.6.6 Cálculo do determinante de uma matriz A

Pelas propriedades do determinante, o determinante não se altera se somarmos um múltiplo de uma linha da matriz à outra, ou seja, se efetuarmos uma combinação linear entre as linhas. Assim, a eliminação de Gauss para se obter uma matriz triangular superior não afeta o valor do determinante

$$\det \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}}_{A} = \det \underbrace{\begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a'_{nn} \end{bmatrix}}_{U}$$

Mas o determinante de uma matriz triangular é o produto dos elementos da matriz. Portanto,

$$\det A = \det U = a'_{11} a'_{22} \ldots h a'_{nn}\,.$$

Atenção: cada operação de pivotamento troca o sinal do determinante! Assim, se para implementar a eliminação de Gauss foram realizados n_p pivotamentos, o determinante será

$$\det A = (-1)^{n_p} a'_{11} a'_{22} \dots a a'_{nn} \,.$$

Exemplo: calcule o determinante da matriz abaixo usando o método de Gauss:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 4 & 3 \\ 6 & 7 & 4 \end{bmatrix} .$$

Usando a eliminação de Gauss sem pivotamento, obtemos

$$U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix} \rightarrow \det A = \det U = 2 \times 2 \times (-1) = -4 \,.$$

Exercício 2: Calcule, usando o método de Gauss com pivotamento parcial a solução do sistema

$$\begin{bmatrix} 4 & 3 & 2 & 2 \\ 2 & 1 & 1 & 2 \\ 2 & 2 & 2 & 4 \\ 6 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 3 \\ 1 \end{bmatrix}$$

5.7 Método de Gauss-Jordan

Este método é uma variante do método de Gauss, onde são eliminados todos os elementos acima e abaixo do pivô. O resultado da eliminação de Gauss-Jordan é uma matriz diagonal. Para ilustrar método, vamos aplicá-lo à solução de um sistema $Ax = b$ e à obtenção simultânea da matriz inversa de A.

Considere a equação matricial

$$A \cdot [x_1 \vee x_2 \vee Y] = [b1 \vee b2 \vee I] \,. \tag{5.4}$$

onde A e Y são matrizes quadradas, x_i e b_i são vetores colunares, I é a matriz identidade, o operador $(\cdot)$ significa um produto de matrizes e o operador $\vee$ significa o aumento de matriz, ou seja, a remoção dos parênteses das matrizes para fazer uma matriz mais larga. É fácil perceber, da equação acima, que os x_1 e x_2 são simplesmente a solução das equações matriciais

$$A \cdot x_1 = b_1 \,,$$
$$A \cdot x_2 = b_2 \,,$$

e que a matriz Y é a inversa de A, ou seja

$$A \cdot Y = I \,.$$

Para simplificar, mas sem perda de generalidade, vamos podemos escrever explicitamente a Eq. (5.4) usando matrizes de ordem 3

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \left\{ \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} \vee \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix} \vee \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} \right\} = \left\{ \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \vee \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \vee \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

Usando a notação de matrizes aumentada, o sistema acima fica

$$\left[\begin{array}{ccc|c|c|ccc} a_{11} & a_{12} & a_{13} & b_{11} & b_{12} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & b_{21} & b_{22} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & b_{31} & b_{32} & 0 & 0 & 1 \end{array}\right]$$

5.7.1 Exemplo de aplicação: inversão de matrizes

$$A \cdot [x \vee Y] = [b \vee I]\,. \tag{5.5}$$

$$\left[\begin{array}{ccc|c|ccc} a_{11} & a_{12} & a_{13} & b_1 & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & b_2 & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & b_3 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c|ccc} 2 & 1 & 1 & 7 & 1 & 0 & 0 \\ 4 & 4 & 3 & 21 & 0 & 1 & 0 \\ 6 & 7 & 4 & 32 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c|ccc} 2 & 1 & 1 & 7 & 1 & 0 & 0 \\ 0 & 2 & 1 & 7 & -2 & 1 & 0 \\ 0 & 4 & 1 & 21 & -3 & 0 & 1 \end{array}\right]$$

A partir daqui, elimina-se os elementos superiores e inferiores ao pivô:

$$\left[\begin{array}{ccc|c|ccc} 2 & 0 & 1/2 & 7/2 & 2 & -1/2 & 0 \\ 0 & 2 & 1 & 7 & -2 & 1 & 0 \\ 0 & 0 & -1 & -3 & 1 & -2 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c|ccc} 2 & 0 & 0 & 2 & 5/2 & -3/2 & 1/2 \\ 0 & 2 & 0 & 4 & -1 & -1 & 1 \\ 0 & 0 & -1 & -3 & 1 & -2 & 1 \end{array}\right]$$

Finalmente, normaliza-se a matriz, de forma que à esquerda ficamos com uma matriz identidade. Obtém-se, assim, o vetor solução do problema e a matriz inversa de A.

$$\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \underbrace{\begin{array}{c} 1 \\ 2 \\ 3 \end{array}}_{\equiv X} \underbrace{\left[\begin{array}{ccc} 5/4 & -3/4 & 1/4 \\ -1/2 & -3/2 & 1/2 \\ -1 & 2 & -1 \end{array}\right]}_{\equiv A^{-1}}\,.$$

5.7.2 Pivotamento total

Vimos acima (seção 5.6.4) um exemplo em que o pivotamento parcial foi usado para evitar erros de arredondamento que podem aparecer quando o multiplicador fica muito pequeno. Partindo do princípio que os erros de arredondamento são tão menores quanto maiores forem os multiplicadores, em geral obtém-se melhores resultados empregando-se o *pivotamento total*, em que se pivotam colunas, além das linhas, de forma a sempre manter o maior termo de uma data linha como pivô.

O pivotalmento total pode ser empregado devido à seguinte propriedade da Álgebra Linear:

Propriedade: a solução de um sistema linear não é alterada quando trocamos de lugar duas colunas i e j de A, *desde que se troquem as duas linhas correspondentes nos vetores x e na matriz Y* (Eq. 5.5).

Desta forma, vemos que as operações de troca de colunas embaralham o(s) vetor(es) das incógnitas e a matriz inversa, de forma que para empregar o pivotamento total devemos manter um registro das trocas de colunas efetuadas para podermos colocar a solução final na ordem correta. Para exemplificar, vamos resolver novamente o sistema da Eq. (5.2). Após a eliminação da primeira coluna, obtemos

$$\left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ 0 & -0.001 & 6 & 6.001 \\ 0 & 2.5 & 5 & 2.5 \end{array}\right] .$$

Vamos agora trocar de lugar as colunas 2 e 3, de forma que o coeficiente 6 será o pivô

$$\left[\begin{array}{ccc|c} 10 & 0 & -7 & 7 \\ 0 & 6 & -0.001 & 6.001 \\ 0 & 5 & 2.5 & 2.5 \end{array}\right] .$$

Eliminando o segundo termo da terceira linha, obtemos

$$\left[\begin{array}{ccc|c} 10 & 0 & -7 & 7 \\ 0 & 6 & -0.001 & 6.001 \\ 0 & 0 & 2.50083 & 2.50083 \end{array}\right] ,$$

que pode ser facilmente resolvido para obtermos $x_p = (0, 1, -1)$. Como fizemos uma troca de colunas, a solução final é obtida trocando-se de lugar as linhas 2 e 3 de x_p, ou seja, $x = (0, -1, 1)$.

5.8 Refinamento da Solução

Seja o sistema

$$Ax = b\,. \tag{5.6}$$

Resolvendo por Gauss ou Gauss-Jordan, obtemos a solução $x^{(0)}$. Sabemos que erros de arredondamento podem ocorrer quando se resolve um sistema linear pelo método de eliminação, podendo comprometer o resultado obtido. Mesmo utilizando pivoteamento total, não se pode assegurar que a solução obtida seja exata.

Inicialmente, notemos que é trivial verificarmos se a solução de um sistema está correta, para isso basta multiplicarmos a matriz A pela solução obtida, $x^{(0)}$, e o resultado deve ser b. Numericamente, esta verificação deve ser feita impondo-se um critério de convergência do tipo:

$$\left|\frac{b_i^{(0)} - b_i}{b_i}\right| < \epsilon, i = 1, \ldots, n\,,$$

onde $b^{(0)}$ é o vetor obtido do produto $Ax^{(0)}$.

O que fazemos quando o resultado obtido $x^{(0)}$ não passa pelo critério de convergência? Uma possibilidade é fazermos o refinamento da solução, como delineado a seguir. Vamos chamar de erro a diferença entre o valor verdadeiro, x, e o valor obtido, $e^{(0)} = x - x^{(0)}$. Substituindo no sistema (5.6), temos

$$A(x^{(0)} + e^{(0)}) = b \tag{5.7}$$

$$Ae^{(0)} = b - Ax^{(0)} \equiv r^{(0)}\,. \tag{5.8}$$

Resolvendo o sistema (5.8) determinamos $e^{(0)}$, a partir do qual podemos fazer uma nova estimativa para a solução: $x^{(1)} = x^{(0)} + e^{(0)}$. Caso $x^{(1)}$ obedeça ao critério de convergência estipulado, teremos encontrado nossa solução. Caso contrário, podemos refinar novamente a solução obtendo uma estimativa para o erro de $x^{(1)}$ resolvendo o sistema

$$Ae^{(1)} = b - Ax^{(1)} \equiv r^{(1)} \,. \tag{5.9}$$

Este processo pode ser executado quantas vezes desejarmos. É fundamental que as operações envolvidas nos cálculos dos resíduos sejam feitas em dupla precisão.

Importante: Como o refinamento envolve a resolução de vários sistemas que compartilham a mesma matriz A, podemos empregar o procedimento descrito na seção 5.6.5 para tornar o processo mais eficiente.

5.9 Sistemas mal-condicionados

Estes sistemas também são conhecidos pelo termo mal-condicionados (*"ill conditioned"*, em inglês). Vejamos um exemplo.

$$\begin{cases} x + \ y = \ 1 \\ 99x + 100y = 99.5 \end{cases}$$

A solução deste sistema é única e exata: $x = 0.5$, $y = 0.5$. Agora considere o sistema

$$\begin{cases} x + \ y = \ 1 \\ 99.4x + 99.9y = 99.2 \end{cases} ,$$

cuja solução única e exata é $x = 1.4$, $y = -0.4$, muito diferente da do sistema anterior, apesar dos coeficientes serem parecidos.

Graficando as retas no plano (x, y) vemos porque isto acontece: as retas correspondentes às equações são quase paralelas, ou seja, as equações são quase linearmente dependentes.

Uma maneira de se "medir" o condicionamento de uma traz é calcular seu determinante. Entretanto, o valor do determinante depende da norma (módulo) de cada um dos seus vetores. Assim, devemos normalizar os vetores para depois calcular o determinante. Isto produzirá um determinante com valor (em módulo) no intervalo $[0, 1]$. Se o módulo do determinante for próximo de zero, então o sistema é mal-condicionado.

Por exemplo, vamos considerar o primeiro exemplo acima. Normalizando os vetores da matriz

$$\begin{bmatrix} 1 & 1 \\ 99 & 100 \end{bmatrix} ,$$

obtemos

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 99/140.716 & 100/140.716 \end{bmatrix} .$$

O módulo do determinante da matriz normalizada é

$$\left| \frac{1}{\sqrt{2}} \frac{100}{140.716} - \frac{1}{\sqrt{2}} \frac{99}{140.716} \right| \approx 5 \times 10^{-3} \,,$$

o que demonstra, quantitativamente, que a matriz original é mal-condicionada.

Há outras medidas do condicionamento de uma matriz, assim como há fórmulas que relacionam o erro cometido no método de Gauss ou Gauss-Jordan com essas medidas e o número de algarismos significativos utilizado. Isto, porém, está além do escopo destas notas. Veja, por exemplo, a seção sobre *singular value decomposition* no Numerical Recipes.

5.10 Decomposição LU

Suponhamos que se possa escrever a matriz A como o produto de duas matrizes

$$A = L \cdot U \,, \tag{5.10}$$

onde L é uma matriz triangular inferior (tem elementos somente na diagonal e abaixo) e U é uma matriz diagonal superior (com elementos somente na diagonal e acima). Para o caso de uma matriz 4×4, Eq. (5.10) ficaria

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \underbrace{\begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix}}_{\substack{L \\ \text{triangular} \\ \text{inferior}}} \underbrace{\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}}_{\substack{U \\ \text{triangular} \\ \text{superior}}} .$$

Pode-se usar esta decomposição para resolver o conjunto de equações lineares

$$Ax = (LU)x = L(Ux) = b \,.$$

Inicialmente resolvemos para o vetor y tal que

$$Ly = b \tag{5.11}$$

e depois resolvemos

$$Ux = y \,, \tag{5.12}$$

para obter a solução final.

Qual a vantagem em quebrarmos um sistema em dois sistemas sucessivos? A vantagem é que a solução de um sistema triangular é trivial. Dessa forma, o sistema (5.11) é resolvido por substituição para frente:

$$y_1 = \frac{b_1}{l_{11}} \tag{5.13}$$

$$y_i = \frac{1}{l_{ii}} \left[b_i - \sum_{j=1}^{i} l_{ij} y_i \right] , \; i = 2, 3, \ldots, n \,, \tag{5.14}$$

e o sistema (5.12) por substituição para trás:

$$x_n = \frac{y_n}{u_{nn}} \tag{5.15}$$

$$x_i = \frac{1}{u_{ii}} \left[y_i - \sum_{j=i+1}^{n} u_{ij} x_j \right] , \; i = n-1, n-2, \ldots, 1 \,. \tag{5.16}$$

5.10.1 Efetuando a Decomposição LU

Como podemos achar L e U dado A? Abaixo vamos delinear um algoritimo bastante utilizado, que pode ser estudado em mais detalhes no Numerical Recipes. Vamos escrever explicitamente o componente $i.j$ da Eq. (5.10). Este componente é sempre uma soma que começa co,

$$l_{i1}u_{1j} + \cdots = a_{ij}\,.$$

O número de termos da soma depende se $i < j$, $i > j$, ou $i = j$. De fato, temos os três casos acima

$$i < j : l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ii}u_{ij} = a_{ij} \tag{5.17}$$

$$i = j : l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ii}u_{jj} = a_{ij} \tag{5.18}$$

$$i > j : l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ij}u_{jj} = a_{ij} \tag{5.19}$$

As Eqs. (5.17) — (5.19) perfazem n^2 equações para $n^2 + n$ incógnitas (note que a diagonal está representada duas vezes). Trata-se, assim, de um sistema indeterminado. Para resolver este sistema, deve-se, assim, *especificar arbitrariamente valores para n incógnitas*. Um procedimento muito usado para resolver a decomposição é o Algoritmo de Crout, que resolve de forma trivial as equações acima para todos os l's e u's simplesmente rearranjando as equações em determinada ordem. O algoritmo é como se segue:

- Faça l_{ii}, $i = 1, \cdots, n$ (de forma a reduzir o número de incógnitas para n^2);
- Para cada $j = 1, \cdots, n$, faça os dois procedimentos seguintes:
  - Primeiramente, para $i = 1, \cdots, j$, use Eqs. (5.17) e (5.18), e a condição acima, para determinar os u_{ij}, ou seja

$$u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}\,. \tag{5.20}$$

    Quando $i = 1$ a soma anterior é tomada como zero.
  - Em segundo lugar, para $i = j+1, \cdots, n$ use (5.17) para achar os l_{ij}, da seguinte maneira

$$l_{ij} = \frac{1}{u_{jj}}\left(a_{ij} - \sum_{k=1}^{j-1} l_{ik}u_{kj}\right)\,. \tag{5.21}$$

    Certifique-se de executar ambos os procedimentos antes de passar para o próximo j!

A chave para compreender o procedimento acima é que os l's e u's que ocorrem no lado direito das equações (5.20) e (5.21) já estão sempre determinados no momento em que são necessários (por isso que o método de Crout é basicamente uma ordem em que as equações devem ser resolvidas). Vemos, também, que cada a_{ij} é usado apenas uma vez durante o processo. Isso significa que os l's e u's podem ser armazenados nos lugares que os termos a's ocupavam na memória do computador. Ou seja, o método de Crout substitui a matriz original A por uma matriz combinada dos elementos de L e U:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ l_{21} & u_{22} & u_{23} & u_{24} \\ l_{31} & l_{32} & u_{33} & u_{34} \\ l_{41} & l_{42} & l_{43} & u_{44} \end{bmatrix}\,.$$

O pivotamento, tal como no caso dos métodos de Gauss e Gauss-Jordan, é essencial para a estabilidade do método de Crout. Quando se emprega o pivotamento parcial, na realidade não se decompõe a matriz A na sua forma LU, mas sim uma permutação das linhas de A. Para ver como efetuar o pivotamento no método de Crout, consulte o capítulo 2 do Numerical Recipes.

Qual a vantagem da Decomposição LU sobre o método de Gauss? Como listado na seção 5.12, o número de operações necessárias para efetuar a decomposição é da ordem de $1/3\,n^3$, exatamente o mesmo número de passos necessários para fazer a eliminação de Gauss. Na literatura, frequentemente cita-se uma vantagem da Decomposição LU que é o fato de que uma vez tendo-se L e U é trivial obter a solução para um número arbitrário de vetores de termos independentes (ou seja, resolve-se facilmente um conjunto de sistema de equações lineares). Entretanto, o mesmo procedimento pode ser feito de forma igualmente eficiente à partir do procedimento delineado na seção 5.6.5.

Conclusão: o método de Gauss e o método da Decomposição LU são igualmente eficientes quando se trata de resolver um sistema de equações lineares, ou um conjunto de sistemas de equações lineares.

5.10.2 Um caso especial: decomposição LU de matrizes tridiagonais

Um caso particular em que a decomposição LU oferece uma solução eficiente é no caso de matrizes tridiagonais. Suponha que A seja uma matriz na forma

$$A = \begin{bmatrix} b_1 & c_1 & & & & \\ a_2 & b_2 & c_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & a_{n-1} & b_{n-1} & c_{n-1} & \\ & & & a_n & b_n & \end{bmatrix},$$

ou seja, apenas os elementos da diagonal principal e das diagonais imediatamente acima e abaixo são não nulos. Neste caso, a decomposição LU então tem uma forma simples

$$L = \begin{bmatrix} 1 & & & & \\ l_2 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & l_n & 1 \end{bmatrix},$$

$$U = \begin{bmatrix} u_1 & v_1 & & & \\ & u_2 & v_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & v_{n-1} \\ & & & & u_n \end{bmatrix}.$$

É fácil demonstrar que a determinação dos l's, u's e v's é feita através da seguintes relações de recorrência:

$$\begin{cases} u_1 = b_1 \\ l_j = a_j/u_{j-1} \\ u_j = b_j - l_j c_{j-1},\ j = 2, 3, \ldots, n \end{cases}$$

Note que nas relações acima está implícito que $v_i = c_i$.

Tendo determinado as matrizes L e U através das relações acima, o procedimento para determinar o vetor solução x do sistema $Ax = d$ (note que aqui usamos d para evitar confusão com os b's da matriz tridiagonal) é simples. Inicialmente calcula-se a solução do sistema $Ly = d$ através da substituição para frente:

$$\begin{cases} y_1 = d_1 \\ y_i = d_i - l_i y_{i-1}, \ i = 2, \ldots, n \end{cases}$$

Calcula-se, então, o vetor solução x resolvendo-se o sistema $Ux = y$ por substituição para trás:

$$\begin{cases} x_n = \frac{y_n}{u_n} \\ x_k = y_k - c_k \frac{x_{k+1}}{u_k}, \ k = n-1, \ldots, 1 \end{cases}$$

O procedimento descrito acima é muito eficiente do ponto de vista computacional e pode ser implementado com facilidade em duas subrotinas, uma para o cálculo da decomposição e outra para a solução do sistema. Note que o fato de que a decomposição LU de uma matriz tridiagonal também é tridiagonal simplifica muito as substituições para frente e para trás. Veremos no Cap. 6 que esta solução para um sistema tridiagonal será muito útil para calcular os coeficientes da interpolação por spline cúbica.

5.11 Forma alternativa para o cálculo da matriz inversa

Denotemos a matriz inversa de A por B, tal que:

$$AB = I \,,$$

onde I é a matriz identidade. Como usar a decomposição LU ou o método descrito na seção (5.6.5) para encontrar B? Isso é feito simplesmente escrevendo a equação matricial acima para cada uma das colunas de B, ou seja,

$$A \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{41} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} ,$$

$$A \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \\ b_{42} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} ,$$

e assim por diante. Ou seja, o cálculo da matriz inversa reduz-se a resolver um conjunto de n sistemas lineares em que os vetores de termos independentes são as diferentes colunas da matriz identidade.

5.12 Comparando Gauss, Gauss-Jordan, e Decomposição LU

Concluímos esta parte do capítulo comparando a eficiência dos três métodos diretos estudados. Temos no quadro abaixo a comparação do número de operações empregadas em cada método para as diferentes tarefas da álgebra linear.

	método	operações
Solução de	Gauss	$1/3\,n^3$
Sistemas	Gauss-Jordan	$1/2\,n^3$
Lineares	LU	$1/3\,n^3$
Inversão	Gauss	$5/6\,n^3$
de	Gauss-Jordan	n^3
Matriz	LU	$5/6\,n^3$
m	Gauss	$\frac{1}{3}n^3 + \frac{1}{2}mn^2$
lados	Gauss-Jordan	$\frac{1}{2}n^3 + mn^2$
direitos	LU	$\frac{1}{3}n^3 + \frac{1}{2}mn^2$

5.13 Métodos Iterativos

Os métodos que vimos até agora (Gauss, Gauss-Jordan, Decomposição LU) são conhecidos como métodos diretos, pois a solução é obtida através da manipulação direta das equações do sistema. Tais métodos podem se tornar ineficientes quando o número de equações fica muito grande ($n \gtrsim 100$), pois o número de operações de ponto-flutuante é $\mathcal{O}(n^3)$. Mais detalhes em Blum(1972), p.131.

Nos métodos ditos iterativos (também chamados de métodos indiretos), arbitra-se um vetor inicial $x^{(0)}$ para a solução e calcula-se uma nova estimativa da solução, $x^{(1)}$ como função de $x^{(0)}$ e assim sucessivamente, ou seja,

$$x^{k+1} = g(x^k)\,,$$

onde k é a k-ésima iteração e g representa uma função qualquer. O processo é repetido até obter a precisão desejada, que se traduz em uma diferença muito pequena entre x^{k+1} e x^k.

Nota: não confunda métodos iterativos com a melhora iterativa da solução, apresentada na seção 5.8.

5.13.1 Método de Jacobi

Seja um sistema linear de ordem n

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = & b_2 \\ \vdots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = & b_n \end{cases}$$

Podemos reescrevê-lo na seguinte forma

$$\begin{cases} x_1 & = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3 - \ldots - a_{1n}x_n) \\ x_2 & = \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3 - \ldots - a_{2n}x_n) \\ \vdots & \\ x_n & = \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - a_{n2}x_2 - \ldots - a_{n,n-1}x_{n-1}) \end{cases}$$

No método de Jacobi, escolhemos arbitrariamente um vetor inicial $x^{(0)}$ e substituímos no lado direito das equações acima obtendo um novo vetor $x^{(1)}$. Repetindo-se o processo k vezes, vemos que a k-ésima estimativa da solução é obtida da seguinte relação de recorrência:

$$\begin{cases} x_1^{(k+1)} & = \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - \ldots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} & = \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k)} - \ldots - a_{2n}x_n^{(k)}) \\ \vdots & \\ x_n^{(k+1)} & = \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k)} - \ldots - a_{n,n-1}x_{n-1}^{(k)}) \end{cases}$$

ou, de forma mais compacta,

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j^{(k)} \right) .$$

Dizemos que o processo iterativo converge se, para a sequência de aproximações gerada, dado $\epsilon > 0$, existir um j tal que para todo $k > j$ e $i = 1, 2, \ldots, n$, $|x_i^k - \bar{x}_i \leq \epsilon$, onde $\bar{x}$ é a solução do sistema. Como na prática não a conhecemos, torna-se necessário um critério de parada para o processo iterativo. Um possível critério é impor que a variação relativa entre duas aproximações consecutivas seja menor que ϵ. Dado $x^{(k+1)}$ e $x^{(k)}$, tal condição é escrita como

$$\max \left\{ \left| \frac{x_i^{(k+1)} - x_i^{(k)}}{x_i^{(k)}} \right|, i = 1, \ldots, n \right\} \leq \epsilon \, . \tag{5.22}$$

Exemplo: considere o seguinte sistema de equações

$$\begin{cases} 4x_1 + 2x_2 + x_3 & = 11 \\ -x_1 + 2x_2 & = 3 \\ 2x_1 + x_2 + 4x_3 & = 16 \end{cases} ,$$

cuja solução é $x = (1, 2, 3)$. Rescrevendo as equações como

$$\begin{cases} x_1 & = \frac{11}{4} - \frac{1}{2}x_2 - \frac{x_3}{4} \\ x_2 & = \frac{3}{2} + \frac{1}{2}x_1 \\ x_3 & = 4 - \frac{1}{2}x_1 - \frac{1}{4}x_2 \end{cases} ,$$

temos que as relações de recorrência, pelo método de Jacobi, são

$$\begin{aligned} x_1^{(k+1)} &= \frac{11}{4} - \frac{1}{2}x_2^{(k)} - \frac{x_3^{(k)}}{4} \, , \\ x_2^{(k+1)} &= \frac{3}{2} + \frac{1}{2}x_1^{(k)} \, , \\ x_3^{(k+1)} &= 4 - \frac{1}{2}x_1^{(k)} - \frac{1}{4}x_2^{(k)} \, . \end{aligned} \tag{5.23}$$

Começando com um vetor arbitrário $x^{(0)} = [1, 1, 1]$ obtemos

$$\begin{aligned} x_1^{(1)} &= \frac{11}{4} - \frac{1}{2} \cdot 1 - \frac{1}{4} \cdot 1 = 2 \\ x_2^{(1)} &= \frac{3}{2} + \frac{1}{2} \cdot 1 = 2 \\ x_3^{(1)} &= 4 - \frac{1}{2}.1 - \frac{1}{4} \cdot 1 = \frac{13}{4} \end{aligned}$$

Substituindo $x^{(1)}$ do lado direito do sistema (5.26) obtemos

$$x_1^{(2)} = \frac{11}{4} - \frac{1}{2}\cdot 2 - \frac{1}{4}\cdot\frac{13}{4} = \frac{15}{16}$$
$$x_2^{(2)} = \frac{3}{2} + \frac{1}{2}\cdot 2 = \frac{5}{2}$$
$$x_3^{(2)} = 4 - \frac{1}{2}.2 - \frac{1}{4}\cdot 2 = \frac{5}{2}$$

Na tabela abaixo listamos os resultados para as 5 primeiras iterações. Vemos que a sequência converge, e atinge uma precisão de aproximadamente 5% em 5 iterações.

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\max\{\lvert(x_i^{(k)} - x_i^{(k-1)})/x_i^{(k)}\rvert, i = 1,\ldots,n\}$
0	1	1	1	-
1	2	2	13/4	9/13
2	15/16	5/2	5/2	3/10
3	7/8	63/62	93/32	17/63
4	133/128	31/16	393/128	7/131
5	519/512	517/256	767/256	21/517

5.13.2 Convergência do Método de Jacobi

(REVISAR)

Vamos escrever a matriz A como

$$A = L + D + U$$

onde L é a *"lower triangular matrix"* (sem diagonal); U *"upper triangular matrix"* (sem diagonal); e D a matriz diagonal.

Desta forma,

$$Ax = (L + D + U)x = b$$
$$Dx = -(L + U)x + b$$
$$x = D^{-1}[-(L + U)x + b]$$
$$x = Jx + c$$

onde $J = -D^{-1}(L + U)$ e $c = D^{-1}b$. Aplicando o método iterativo teremos

$$x^{(k+1)} = Jx^{(k)} + c, \text{ onde } J = -\begin{bmatrix} 0 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} & \cdots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & \frac{a_{23}}{a_{22}} & \cdots & \frac{a_{2n}}{a_{22}} \\ \vdots & & 0 & & \vdots \\ \vdots & & & 0 & \vdots \\ \frac{a_{n1}}{a_{nn}} & \frac{a_{n2}}{a_{nn}} & \cdots & \frac{a_{n,n-1}}{a_{nn}} & 0 \end{bmatrix}$$

Partindo de $x^{(0)}$ e fazendo sucessivamente a iteração temos

$$x^{(k)} = \underbrace{J^k}_{\substack{\text{elevado} \\ \text{a } k}} x^{(0)} + [1 + J + J^2 + \ldots + J^{k-1}]c \tag{5.24}$$

Para que convirja, requer que

$$\lim_{k\to\infty} J^k = [0]$$

O que implica que $\lim_{k\to\infty}[1+J+J^2+\ldots+J^{k-1}] = (1-J)^{-1}$. Assim quando (5.24) é satisfeita, $x = \lim_{k\to\infty} x^{(k)}$ existe e $x = 0+(1-J)^{-1}c$, isto é, $(1-J)x = c$ ou $x = Jx+c$.

Mas a condição (5.24) é válida se e somente se todos os auto valores da matriz J forem em módulo < 1.

Seja $\rho_s = \max|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|$ onde $|\lambda_i|$ são os autovalores da matrix J. ρ_s é também chamado de raio espectral (*"spectral radius"*).

Então para atingir precisão p após k iterações devemos ter

$$\rho_s^k \approx 10^{-p} \;\rightarrow\; \boxed{k \approx -\frac{p\ln 10}{\ln \rho_s}}$$

Assim se ρ_s estiver próximo de 1 a convergência será muito lenta. Existem métodos de aceleração. Ver *Quinney* e *NR* seção 19.5.

Determinar os auto valores da matriz J requerirá outro algoritmo, em geral. Na prática, muitas vezes é mais fácil testar numericamente a convergência.

Critério das linhas

Uma condição mais simples de convergência , porém apenas suficiente, é que o sistema possua diagonal principal estritamente dominante, ou seja,

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}} |a_{ij}|, i = 1, \ldots, n \tag{5.25}$$

que é chamado de critério das linhas. Note que por este critério, os elementos da diagonal principal nunca podem ser nulos.

Exercício: mostre que a matriz do sistema

$$\begin{bmatrix} 4 & 2 & 1 \\ -1 & 2 & 0 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 11 \\ 3 \\ 16 \end{bmatrix}$$

satisfaz o critério das linhas.

Por ser um critério apenas suficiente, um sistema que não satisfaz o critério das linhas pode convergir. Além disso, alterando a ordem das linhas ou colunas pode-se tornar um sistema convergente em divergente e vice-versa.

5.13.3 Método de Gauss-Seidel

O método de Gauss-Seidel é muito semelhante ao método de Jacobi, mas em geral apresenta uma convergência mais rápida. Neste método, aproveita-se os valores já calculados em uma iteração (ex:, $x_1^{(k+1)}$) para a estimativa dos termos seguintes.

As relações de recorrência tomam a seguinte forma

$$\begin{cases} x_1^{(k+1)} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \ldots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \ldots - a_{2n}x_n^{(k)}) \\ x_3^{(k+1)} &= \frac{1}{a_{22}}(b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} - \ldots - a_{2n}x_n^{(k)}) \\ \vdots & \\ x_n^{(k+1)} &= \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k+1)} - \ldots - a_{n,n-1}x_{n-1}^{(k+1)}) \end{cases}$$

ou, de forma mais compacta,

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} \right) .$$

Exemplo: vamos considerar novamente o sistema

$$\begin{cases} 4x_1 + 2x_2 + x_3 &= 11 \\ -x_1 + 2x_2 &= 3 \\ 2x_1 + x_2 + 4x_3 &= 16 \end{cases} .$$

As relações de recorrência, pelo método de Gauss-Seidel, são

$$\begin{aligned} x_1^{(k+1)} &= \frac{11}{4} - \frac{1}{2} x_2^{(k)} - \frac{1}{4} x_3^{(k)} , \\ x_2^{(k+1)} &= \frac{3}{2} + \frac{1}{2} x_1^{(k+1)} , \\ x_3^{(k+1)} &= 4 - \frac{1}{2} x_1^{(k+1)} - \frac{1}{4} x_2^{(k+1)} . \end{aligned} \tag{5.26}$$

Começando novamente com o vetor $x^{(0)} = [1, 1, 1]$ obtemos sucessivamente

$$x^{(1)} = \begin{pmatrix} 2 \\ 5/2 \\ 19/8 \end{pmatrix} , \; x^{(2)} = \begin{pmatrix} 29/32 \\ 125/64 \\ 783/256 \end{pmatrix} , \; x^{(3)} = \begin{pmatrix} 1033/1024 \\ 4095/2048 \\ 24541/8192 \end{pmatrix} \approx \begin{pmatrix} 1.0087 \\ 1.9995 \\ 2.9957 \end{pmatrix} .$$

Note que, neste exemplo, a taxa de convergência é muito maior.

5.13.4 Convergência do Método de Gauss-Seidel

O critério das linhas também pode ser aplicado ao método de Gauss-Seidel, mas, como no método de Jacobi, trata-se apenas de uma condição suficiente.

Para o método de Gauss-Seidel existe um outro critério, menos restritivo que o critério das linhas, chamado *critério de Sassenfeld*. Seja

$$M = \max_{1 \le i \le n} \beta_i ,$$

onde os β_i são definidos por

$$\begin{aligned} \beta_1 &= \frac{|a_{12}| + |a_{13}| + \cdots + |a_{1n}|}{|a_{11}|} , \\ \beta_i &= \frac{\sum_{j=1}^{i-1} \beta_j |a_{ij}| + \sum_{j=i+1}^{n} |a_{ij}|}{|a_{ii}|} . \end{aligned}$$

A condição $M < 1$ é suficiente para que as aproximações sucessivas pelo método de Gauss-Seidel convirjam.

Muito importante: A convergência (ou não) dos métodos de Jacobi ou Gauss-Seidel independe do vetor inicial escolhido.

Exercício: use o método de Gauss-Seidel para resolver o sistema abaixo. Verifique se a matriz satisfaz o critério das linhas e o critério da Sassenfeld.

$$\begin{bmatrix} 10 & -2 & -2 & 1 \\ -2 & 5 & -1 & -1 \\ 1 & 1/2 & -6 & 1 \\ -1 & -1 & 0 & 20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ -9 \\ 17 \end{bmatrix}$$